

- Distintos tipos de anotación según el vocabulario utilizado

**Basada en Dublin Core**

The *contributor* and *creator* is the flight booking service "www.flightbookings.com".  
The *date* would be January 1<sup>st</sup>, 2003, in case that the HTML page has been generated on that specific date.  
The *description* would be something like "flight details for a travel between Madrid and Seattle via Chicago on February 8th, 2004".  
The document *format* is "HTML".  
The document *language* is "en", which stands for English

**Flight details**

Outbound	Flight details
Leaving from <b>Madrid</b> - Barajas - Spain on Saturday 08 February 2003 at <b>11:50</b> Arriving in <b>Chicago</b> - O'Hare International - United States of America same day at <b>14:10</b> Airline: American Airlines Flight No. AA 7615 Type of aircraft: Airbus Industrie A340 All Series PAVH	Leaving from <b>Chicago</b> - O'Hare International - United States of America on Saturday 08 February 2003 at <b>16:48</b> Arriving in <b>Seattle</b> - Seattle/Tacoma International - United States of America same day at <b>19:23</b> Airline: American Airlines Flight No. AA 1605 Type of aircraft: non referenced/B

**Basada en tesauros**

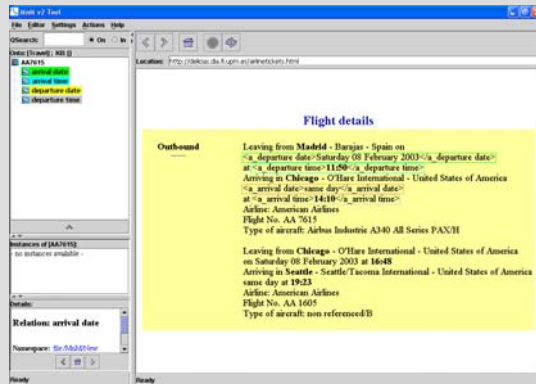
Madrid is a reference to the term with ID 7010413 in the thesaurus, which refers to the city of Madrid in Spain.  
Spain is a reference to the term with ID 1000095, which refers to the kingdom of Spain in Europe.  
Chicago is a reference to the term with ID 7013596, which refers to the city of Chicago in Illinois, US.  
United States of America is a reference to the term "United States" with ID 7012149, which refers to the US nation.  
Seattle is a reference to the term with ID 7014494, which refers to the city of Seattle in Washington, US.

**Basada en ontologías**

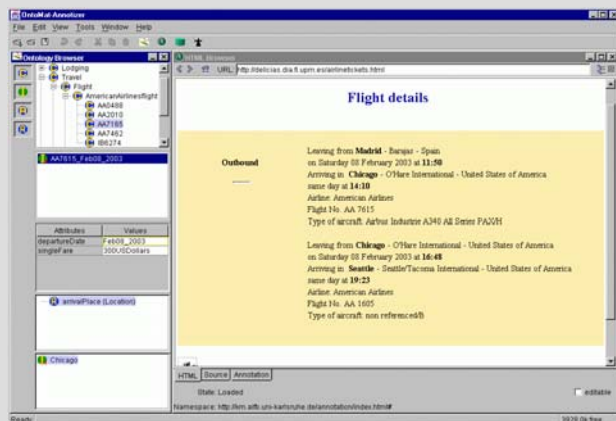
**Concept instances** relate a part of the document to one or several concepts in an ontology. For example, "Flight details" may represent an instance of the concept `Flight`, and can be named as `AA7615_Feb08_2003`, although concept instances do not necessarily have a name.  
**Attribute values** relate a concept instance with part of the document, which is the value of one of its attributes. For example, "American Airlines" can be the value of the attribute `companyName`.  
**Relation instances** that relate two concept instances by some domain-specific relation. For example, the flight `AA7615_Feb08_2003` and the location `Madrid` can be connected by the relation `departurePlace`

- Fuentes
  - Tipos de fuentes
    - Texto: HTML, XML, PDF, etc.
    - Multimedia: imágenes, video, audio, etc.
    - Servicios Web
  - Origen de la fuente de datos
    - Estático: ficheros
    - Dinámico: bases de datos y formularios
- Tecnologías utilizadas
  - Knowledge extraction
    - NLP, IE, Layout
  - Wrapper generation
    - Toolkits, ML, Browsing
- Proceso de anotación
  - Mantenimiento (adaptabilidad a los cambios en la fuente)
    - Verificación
    - Robustez
    - Auto-adaptabilidad
  - Supervisión de las anotaciones
    - Manual
    - Supervisado (semi-automático)
    - No supervisado (automático)

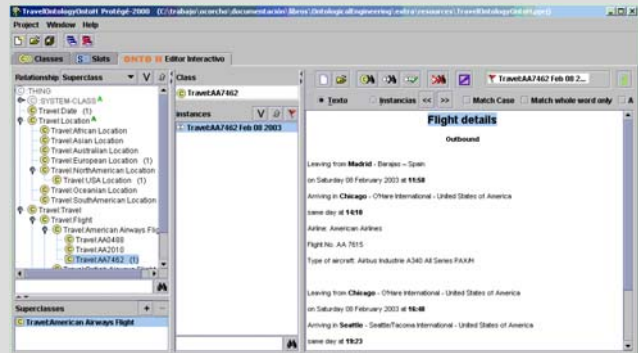
- Aplicación standalone que integra un navegador Web y un visor de ontologías
- Anotación manual con drag&drop
- Anotación semi-automática y automática utilizando herramientas de extracción de información (Amilcare)
- Anotaciones generadas en OCML, RDF y XML



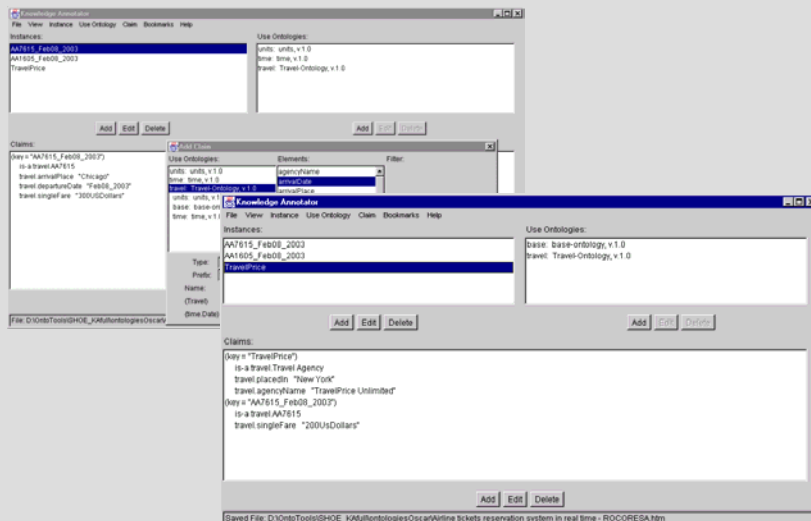
- Aplicación standalone que integra un navegador Web y un visor de ontologías
- Anotación manual con drag&drop
- Anotaciones generadas en RDF y OWL



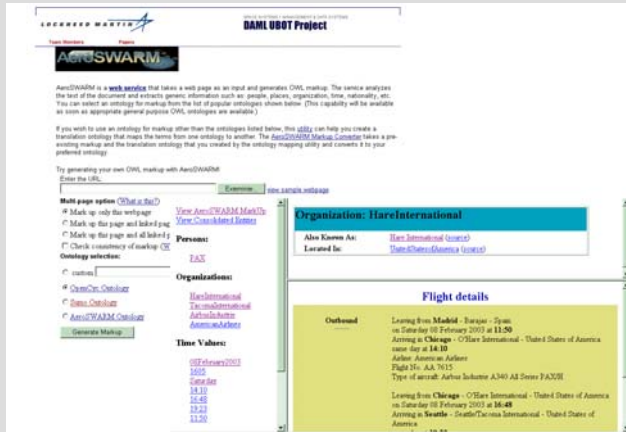
- Plug-in de Protégé que integra un visor de documentos RTF y el visor de ontologías por defecto
- Anotación manual con drag&drop
- Anotación en cascada de conceptos relacionados, mediante reglas de anotación declarativas
- Anotaciones generadas en el formato de Protégé, exportable a diversos lenguajes



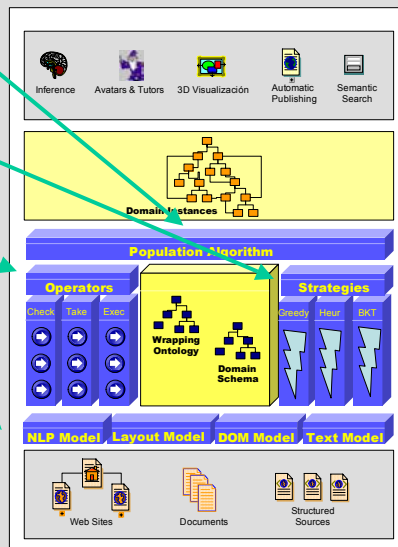
- Aplicación standalone que no contiene navegador Web
- Anotación manual
- Anotaciones generadas en SHOE



- Servidor Web de anotación de documentos Web disponibles en una URL
- Anotación automática, basada en un conjunto de ontologías predefinidas: OpenCyc, SUMO y AeroSWARM
- Anotaciones generadas en RDF



- **Populación** basada en la semántica
  - Conocimiento de wrapping explícito
  - Bootstrapping
- **Estrategias**
  - Heurística
  - Backtracking
  - Fuerza bruta, etc.
- **Diferentes operadores**
  - In-Row
  - Is-Proper-Name
  - Is-Integer-Greater-Than, etc.
- **Diferentes pre-procesados**
  - Lenguaje natural
  - Layout
  - XML/DOM
  - Texto plano
- Sectores en los que se ha aplicado
  - Cultural
  - Subvenciones
  - Financiero
  - Relaciones internacionales



- Corcho O (2005) *Ontology based document annotation: trends and open research problems*. International Journal of Metadata, Semantics and Ontologies 1(1)
- Gómez-Pérez A, Fernández-López M, Corcho O (2004) *Ontological Engineering*. Springer-Verlag
- Handschuh S, Staab S (2003) *Annotation for the Semantic Web*. IOS Press